Preliminary Analytical Considerations In Designing A Terrorism And Extremism Online Network Extractor

Martin Bouchard, PhD; Kila Joffres, MA; Richard Frank, PhD

International CyberCrime Research Center
School of Criminology
Simon Fraser University
mbouchard@sfu.ca,kja4@sfu.ca,rfrank@sfu.ca

Abstract. It is now widely understood that extremists use the Internet in attempts to accomplish many of their objectives. In this chapter we present a web-crawler called the Terrorism and Extremism Network Extractor (TENE), designed to gather information about extremist activities on the Internet. In particular, this chapter will focus on how TENE may help differentiate terrorist websites from anti-terrorist websites by analyzing the context around the use of predetermined keywords found within the text of the webpage. We illustrate our strategy through a content analysis of four types of web-sites. One is a popular white supremacist website, another is a jihadist website, the third one is a terrorism-related news website, and the last one is an official counterterrorist website. To explore differences between these websites, the presence of, and context around 33 keywords was examined on both websites. It was found that certain words appear more often on one type of website than the other, and this may potentially serve as a good method for differentiating between terrorist websites and ones that simply refer to terrorist activities. For example, words such as "terrorist," "security," "mission," "intelligence," and "report," all appeared with much greater frequency on the counterterrorist website than the white supremacist or the jihadist websites. In addition, the white supremacist and the jihadist websites used words such as "destroy," "kill," and "attack" in a specific context: not to describe their activities or their members, but to portray themselves as victims. The future developments of TENE are discussed.

Keywords: Web-Crawler, Extremism, Terrorism, Internet

1 Introduction

Much like any other groups and organizations, extremist and terrorist groups can be found on the Internet, including many who have their official website [21]. Conway [4] has suggested that extremists use the Internet in five general ways: information recruitment, networking, information provision, financing, and recruitment (see also [16]). The Internet's appeal follows from its ability to provide

a broad reach, to provide low costs, to be timely and efficient, and to provide some degree of security and anonymity [8]. Tsfati and Weimann [19] further emphasize that the Internet is extremely well suited to terrorists for the purposes of communication as it is decentralized, uncensored, largely free of control or restrictions, and allows for worldwide access for current members or potential recruits.

Despite the identification of the Internet as a tool for terrorist groups, limited empirical research has been conducted into the nature of the terrorism-related content online. However, researchers involved with the Dark Web Project have started to build knowledge on the content and structure of websites hosting terrorism-related content [1]. For example, Zhou et al. [24] proposed a semiautomated methodology (combining the efficiency of automatic data collection and the accuracy of manual collection) for identifying, classifying, and organizing extremist Website data. The Dark Web Project has since spawned several research studies on the nature of terrorist use of the Internet and their online networks [1]. These methodologies allow for the analysis of extremist Website content, which can further inform counterterrorism research and policy (e.g., in terms of disrupting online networks and intelligence-gathering on terrorist operations).

Information on terrorist communications, ideologies, activities, relationships, and so on, can be produced by studying extremist web-sites. As such, the main objective of this chapter is to lay the foundation for the development of the Terrorism and Extremism Network Extractor (TENE). In its fully developed form, TENE will be a custom-written computer-program that automatically browses the World Wide Web for terrorism and extremism content, collecting information about the pages it visits.

The foundation of this approach is based on a combination of the ground breaking work associated with the Dark Web Project [1] and our previous efforts to study online child exploitation websites [10] [22] [6]. For instance, the Dark Web project has spawned various programs designed to examine particular types of online extremist content (e.g., forums), to visualize and determine the topology of networks of extremist webpages, and to study certain aspect of the content on extremist websites (e.g., technical sophistication, media richness, and web interactivity) [1]. However, an "intelligent" version of TENE that would browse the Web to automatically identify extremism content has yet to be created. Methods to allow TENE the capability to make decisions automatically can be done, for example, by borrowing techniques from the domains of datamining [9] [11] or machine learning [15] [18]. Features of webpages to use for analysis could include keywords, page structure and the location of that page within the Internet. While analyzing the frequency and occurrences of certain keywords on a website may assist in suggesting something about the nature of the content, it is unlikely to be enough for accurately labeling a website as containing and spreading extremist views. A news website, for example, may contain the same words with similar frequencies. The same may potentially be said of a governmental website focused on issues associated with terrorism and counter-terrorism. Unfortunately, the type and content of websites containing

material that can be qualified as "extremism" has yet to be fully understood (the aphorism that one man's terrorist is another man's freedom fighter has yet to find a suitable empirical solution for such purposes). Thus, before undertaking the process of making TENE more intelligent, additional baseline data is needed about the structure and content of single websites in order to establish the ground rules necessary to create a valid web-crawler program.

The current study contributes to this end goal. In this chapter, we use a preliminary version of the web-crawler designed to collect the entire information contained on a single website. In particular, this chapter will focus on how TENE may help differentiate terrorist/extremist websites from other types of related websites by analyzing the context around the use of predetermined keywords found within the HTML of the webpage. We illustrate our strategy through a content analysis of four types of websites. One is a popular white supremacist website, another is a jihadist website, the third one is a terrorism-related news website, and the last one is an official counterterrorist website.

2 Methodology

TENE is a web-crawler that emerges from previous work on extracting online child pornography networks (see [10] [22] [6]). TENE operates by starting the crawling process at user-specified webpages, retrieving the pages from the Internet, analyzing them, and recursively following the links out of the pages. For the purpose of this chapter, the web-crawler starts at a page that covers material broadly associated with extremism or terrorism. Such a webpage can be found by the user, given to the web-crawler by the police, or obtained from terrorism-related literature. The starting website is then retrieved for the crawler, but there is no need to display the content in a web-browser and hence only the HTML (Hypertext Markup Language) of the webpage is retrieved. Certain statistics about the content of webpages are recorded, such as the frequency of user-specified keywords and count of images or videos. In its mature form, TENE will also follow the links found on a webpage if these links point to a webpage that contains extremism or terrorism material. These links will be subsequently explored recursively until certain criteria are met.

As the Internet is extremely large and a crawler would most likely never stop crawling, three conditional limits can be implemented into the web-crawler. These conditions help keep the crawling process under control and the network content-relevant. First, to keep the network extraction time bounded, a limit can be put on the number of pages retrieved (in our previous work on child pornography, that limit was 250,000). Second, the network size may be fixed at a specific number of websites (for example, 500). The webpages are retrieved in such a way that each website is sampled equally, or as equally as possible. Finally, in order to provide some boundaries for the crawl and guide the network extraction process to a relevant network, a set of keywords needs to be defined. For the crawler to include a given webpage in the analysis, the page has to contain a user-defined number of unique keywords.

The end result of the crawling process is knowledge about a set of webservers, including the webpages contained within them, and the links between the webpages. These results are then aggregated up to the server level, with the resulting network summarizing the content on each of the servers, count of keywords, videos, and images, and the links between each of the servers. This essentially creates a map of a terrorism network from the Internet. Note that the version of TENE used for the purpose of this chapter remains within the realm of the initial user-specified website from which it starts. This work will eventually lead to the establishment of rules allowing for automatic identification of a terrorism/extremism-related website from another.

2.1 Keywords

As previously mentioned, to keep the websites crawled in TENE topic-relevant, keywords are used as an inclusion criterion. Previously, with the child exploitation application, specific keywords and their frequencies were used to establish it as a child exploitation website. The keywords were derived from manual analysis of the textual content on child exploitation pages, as well as through contacts with law enforcement. A new set of keywords is derived in this chapter from the terrorism and extremism domain, and includes words that extremist groups are known to include on webpages, such as bomb, recruit, attack, or target (full list can be found in Table 2). A rate of use per webpage will be calculated for each keyword, on each website, in order to detect whether a type of website is more likely to use a word compared to another. The root word method is used, so that "bomb" and "bombing," for example, is meant to be the same word.

Counting the number of instances a keyword was used on a webpage was sufficient in the child exploitation domain because the analysis showed that the keywords accurately indicated the presence of child exploitation, and were very poor at detecting the content of websites dedicated to countering child exploitation. This is expected to be more of a challenge in the terrorism domain, where words such as bomb or attack can be used, for example, by law enforcement-related websites focused on counter-terrorism. As such, TENE has been extended to not only count the number of instances of a keyword, but also capture the circumstances within which the keyword appears. That is, TENE is designed to be content-aware and at present, extracts a minimum of 200 characters before and after a given keyword. This number of characters was determined to be adequate for ascertaining the context of the keywords within their sentences, and with regard to immediately preceding and following sentences. This will allow for an analysis of each keyword in its original context to better understand when and how it is used. It should be noted that while TENE is programmed to extract the context, the analysis of this context is performed manually. This step is expected to be replaced by some text-classification method(s) in TENE's final form.

Table 1. A Description of the Four Websites Selected for Analysis

Website	**# of Pages**	**Host**	**Category**	**Description**
http://jorgevinhedo.sites.uol.com.br	9	The Lashkar-e Tayyiba, a prominent militant jihadist organizations in South Asia.	jihadist website [2] [23]	Posts information about Islam, attacks against Muslims, and includes links to Islamic newsletters and other relevant sources of information.
http://www.natall.com	47	The National Alliance, a white supremacist and white nationalist political organization.	White supremacist website [2] [23]	Includes online publications, broadcasts, a forum, and links to similar websites.
http://www.counterpunch.org	164	Alexander Cockburn and Jeffrey St. Clair, editors of the website.	News website	An American website that posts political news articles, updated every weekday.
http://www.nctc.gov	15	The U.S. government. NCTC integrates and analyzes intelligence pertaining to terrorism, keeps a knowledge bank on terrorism information, and provides support to counterterrorism activities	Counterterrorist website	Includes information about NCTC, its mission, its goals, its products, and its activities. It also provides information on key partners, its director, and a page for children.

2.2 Websites Selected for Analysis

Four different websites were explored using TENE, including a jihadist website, a white supremacist website, a news website, and a counterterrorist website. The jihadist and the white supremacist websites were obtained from research articles emerging from the Dark Web Project, and were labeled as "extremist" or "terrorist" websites by the researchers (see [2] [23]). The counterterrorist website was an official American government website while the news website was automatically identified by the web-crawler in a test run as having many of the keywords input into the program. Table 1 provides a brief description of each website, the number of pages crawled, and the website hosts. The jihadist website was quite small (9 pages in total), with the main page containing most of the information, supplemented by material from the other 8 webpages within the website. The largest website (the news website) contained 164 pages of largely article archives.

3 Results

TENE crawled the webpages within the four aforementioned websites and recorded the number of keywords during this process. The keywords appearing on these websites and their context was explored in order to determine a) whether these websites can be differentiated by the keywords that appear on them and b) whether certain keywords are used in a similar or different manner across websites. Overall, our results show that certain keywords appear to be associated with specific websites. In addition, different websites use the same words in different context and occasionally use different words in the same context.

Table 2 shows the number of keywords mentioned per page; it can be seen that each website has keywords that tend to be specific to it. That is, certain keywords appear more frequently for each website compared to others. For instance, words such as "Allah", "attack", "Islam", "infidel", and "Jihad" occurred more often in the jihadist website than in any other website; these words ranged from 0.13 ("infidel") to 7.38 ("Islam") mentions per page. In contrast, the white supremacist website employed the words "Jew" (13.38 mentions per page) and "white" (20.98 mentions per page) considerably more often than other websites. The counterterrorist website also used specific words at a greater frequency, including "counterterrorism", "intelligence", "terrorist", and "security", and "combat". Conversely, words such as "dead", "kill", "foreign", and "Obama", appeared more commonly on the news website. This suggests that particular keywords may be used to distinguish between types of websites.

This is represented more clearly in Table 3, where individual web-sites and combinations of websites are associated with particular keywords. This table was constructed by comparing the rates of words appearing for each website; if a word for a particular website occurred at a rate of less than one-third the rate of the highest incidence of that word, then no association between that word and the website was marked. This allowed us to classify words that were solely used by one of the four types of websites analyzed, as well as words that were

Table 2. Number of Keywords per Page for each Website

Keyword	**Jihadist**	**White Supremacist**	**News**	**Counterterrorist**
Allah	3.63	0.04	1.07	0
Attack	6.38	0.89	0.87	0.6
Black	0.86	3.11	4.79	0
Bomb	0	0.28	0.8	0.53
Combat	0	0.06	0.1	0.47
Counterterrorism	0	0	0.02	4.33
Dead	0	0.15	1.60	0.07
Destroy	0.38	0.57	0.25	0
East	0	1.45	1.24	0
Fight	0	0.68	0.55	0.07
Foreign	0	0.19	0.59	0
Free Speech	0.5	0.04	0.40	0
Infidel	0.13	0.04	0	0
Intelligence	0	0.15	0.32	7.67
Islam	7.38	0.09	0.55	0
Jew	3.5	13.38	2.37	0
Jihad	2.25	0	0.06	0
Join	0.25	0.70	0.29	0.53
Kill	0.13	0.38	1.39	0.07
Live	0.25	1.55	0.91	0.07
Member	0.13	2.89	1.32	1.27
Mission	0	0.13	0.69	1.6
Obama	0	0.28	5.62	0.13
Race	0.38	19.28	11.12	0.67
Report	1	0.77	2.40	5.8
Security	0	0.17	1.04	4.07
Struggle	0.13	0.45	0.21	0
Terrorist	3.13	0.36	0.6	10.93
Train	0.25	0	0.39	0
Victim	0	0.15	0.26	0
Violence	0.63	0.38	0.55	0.13
West	1.5	1.72	1.99	0
White	0	20.98	0.70	5.27

Table 3. Keywords most strongly associated with specific websites

Website(s)	**J**	**WS**	**N**	**C**	**J & N**
Keywords	Allah	Jew	Dead	Combat	Free Speech
	Attack	White	Foreign	Counterterrorism	Train
	Infidel		Kill	Intelligence	
	Islam		Obama	Security	
	Jihad			Terrorist	
Website(s)	**WS & S**	**N & C**	**WS, N & C**	**J, WS & N**	**J, WS, N & C**
Keywords	Black	Mission	Bomb	Destroy	Join
	East	Report	Member	Violence	
	Fight			West	
	Live				
	Race				
	Struggle				
	Victim				

J = Jihadist; WS = White Supremacist; N = News; C = Counterterrorism

used by various combinations of two or three web-sites. For instance, both the jihadist and the news website used the words "free speech" and "train" more frequently than other web-sites. Similarly, the white supremacist and the news website shared several words, including "black," "East," "fight," "live," "race," struggle," and "victim." The news and counterterrorist website employed the terms "mission" and "report" more frequently than other websites. In some instances, all but one website used a particular word at similar rates. For example, the white supremacist website, the news website, and the counterterrorist website shared the words "bomb" and "member." The jihadist website, the white supremacist websites, and the news websites each used "destroy," "violence," and "west" at comparable rates. Note that the word "join" was similarly popular across all websites.

The type of keywords that appeared more frequently for each web-site followed intuitively from the nature of the website. For a jihadist website, words such as "Allah" and "Jihad" are more relevant to the subject matter of the website. While terms such as "white" and "Jew" may also be used in a jihadist website, these words are more likely to appear in the many discussions of "race" in the white supremacist website. However, both websites also use certain words at similar rates, including "destroy," "violence," and "West", and as will be seen, the websites tend to use such words in a similar context. The content of counterterrorist websites is very different, and the associated keywords reflect its focus on terrorism incidents and intelligence reports. For a news website, articles on deaths and politics are likely to be an important part of the website, and as such, are more popular words. In addition, both counterterrorist and news websites shared the use of "mission" and "report," which further differentiate their content from the more religious and racial focus of the jihadist and white supremacist websites. In these ways, the type of website tends to be represented

in specific words that appear with greater frequency on the particular website compared to others.

4 Contextual Analysis

An initial crawl of the different websites also allowed for the examination of how certain keywords are used within each website. For example, certain words were used by the jihadist and white supremacist websites with similar frequency, but in varying contexts. Words such as "West," which were used with similar frequency (1.5 times per page for the jihadist extremist website and 1.72 times for the white supremacist website), took on very different meanings between the websites. In the jihadist website, "West" was used when discussing:

- "justification for Jihad against US and its terrorist western allies,"
- "anti-Islam sentiments all over the West," and
- "any invader in Afghanistan or Pakistan - be it West, India or Israel - would have to face the collective strength of the Muslims."

In essence, anti-West sentiments were prominent within the website. In contrast, the white supremacist website took pride in the West, discussing concerns about "alien groups" taking over "both in terms of culture and [white] genetic future," expressing concerns over "the decline of the West generally" with its growing multiculturalism, and posting assurances that the "National Alliance will lead [white] people to a secure homeland here in the Western hemisphere."

Other words were also used commonly between the various web-sites, but in different contexts. For instance, both the white supremacist and the news website used the word "black" more often than other websites (3.11 times per page for the white supremacist web-site and 4.79 times per page for the news website compared to 0.86 times per page for the jihadist website and 0 times for the counter-terrorist website). However, the news website used the word as a colour or adjective. In contrast, the white supremacist website largely employed the word to refer to African Americans, doing so in a typically disparaging manner (e.g., by discouraging interracial marriages, criticizing sympathetic portrayals of "Blacks" in the media and images of "Whites" and "Blacks" together, discussing the "problem" of "Black crime," and arguing that "Black History month is destroying the past").

While the same words were sometimes used differently between websites, other words were used to express the same point. For instance, the jihadist and white supremacist websites both used certain words to emphasize the victimization of the groups whose interests they claim to represent. The word "attack," which appeared most frequently on the jihadist website (6.38 times per page) was used to describe attacks against Afghanistan as "terrorism" and to raise awareness about attacks against the Quran and mosques. The word "destroy" (used 0.38 times per page) was employed in a similar manner, with reference to power stations and villages being destroyed. Even the word "Islam" (occurring 7.38 times per page) was often used in a manner that portrayed Muslims to be under attack; for instance:

- "The western and American print and electronic media are continuously spitting venom against Islam,"
- "The Muslims have already suffered too much of violence and tyranny but now the non-Muslim world plans to eliminate the Muslims once for all simply because strong Islam is something intolerable for the non-Muslim forces," and
- "The enemy has already declared a war against Islam".

In addition, the word "kill" (used 0.13 times per page) was further used to emphasize "the killing of innocent people" at the hands of the U.S.

Similarly, when the white supremacist website used words such as "attack," "bomb," "dead" and "destroy," it was to emphasize the victimization of "whites" at the hands of other "races". For example, "attack" (appearing 0.89 times per page) was used in the context of attacks on freedom of speech, various terrorist attacks against Americans, non-white immigrants as "attacks on freedom," and so on. The word "destroy" (occurring 0.57 times per page) was often used in the same way, describing how:

- "Mexicans will destroy America,"
- "Jewish heritage week will destroy American heritage," and
- "multiculturalist movement will destroy the fabric of White America."

The term "fight" (used 0.68 times per page) was also employed in a similar manner, with the white supremacist website noting that white individuals must "fight for the security and survival of [their] people," fight against "white racism," and fight against organized crime by fighting multiculturalism. Finally, the word "kill" was used to discuss various killings of "Whites" by "Blacks," although it was also used it in other contexts, including engaging in Holocaust denial by questioning the killing of Jewish people and discussing the killing of Saddam Hussein. Overall, a tendency emerged for these websites to use certain words in ways that emphasizes their role as victims with a sympathetic cause rather than as aggressors with a violent agenda. This allows websites to set the stage for encouraging action, further propaganda, and/or for recruitment purposes.

Some of the same words were also used from vastly different perspectives. For instance, each of the websites used the word "terrorist" to refer to American activities in the Middle East, but approached the issue from a different angle. Within the jihadist website, the word was used in reference to "U.S. terrorist attacks" against Islamic nations, to describe the "terrorist war against Muslim ummah ["community"]" launched by the U.S., to describe the "U.S. and its western allies' attack on Afghanistan as the worst kind of terrorism," and to describe America's allies as "terrorists." At the same time, the website seeks to dissociate the word terrorism from Muslims, by emphasizing that the "Noble Quran" does not preach terrorism and that mosques are not training grounds for terrorists, despite western "propaganda" to this effect.

The white-supremacist also uses the word "terrorism" largely to describe America's actions in the Middle East, criticizing the U.S. policy surrounding the

"War against Terrorism." For instance, the website states that "Covert Operations are a huge part of the CIA [and] are simply state- sponsored terrorism," further arguing that "The only way we can end our War against Terrorism, is to end the US practice of conducing Terrorism under the guise of Covert Ops." The website also criticizes the use of detention centers such as Guantanamo Bay to hold so-called "terrorists" or suspected terrorists". The general consensus appears to be that the U.S. should focus more on the state of its nation within its borders rather than outside. As such, while both websites condemn American actions in the Middle East, they do so from different stand points and for different reasons.

Overall, the different types of websites can be differentiated by the different frequency of keywords used. However, both jihadist and white supremacist websites use various words for the same purpose (e.g., portraying themselves as victims) and use the same words for different purposes (e.g., attacking or defending the West or for recruitment or general discussion).

5 Discussion

The "National Strategy for Homeland Security" report in the U.S. emphasized that science and technology were important counter-terrorist tools [14]. It has been suggested that the use of information technology will increase national safety [13] by assisting in intelligence gathering through the collection and analysis of terrorism-related data [3]. This renders the creation of web-crawlers designed to detect and extract networks of extremist or terrorist web-sites a valuable enterprise, as these can build knowledge related to the content (i.e., group activities, recruitment processes, propaganda materials, etc.) of such websites and the affiliations of these groups.

This exploratory study used TENE, a specially designed web-crawler, to explore certain content aspects of two extremist websites (a white supremacist website and a jihadist website), with comparisons made to non-extremist websites (a counterterrorist website and a news website). It was found that all websites could be identified by certain keywords; for instance, the jihadist website used words such as "Allah," "Islam," and "Attack" at greater rates than the other websites. In addition, the extremist websites used language in specific ways, with words such as "attack," "destroy," and "dead" being used to emphasize the group's role as a victim. It should be stressed that the number of websites selected is small; as such, this project is entirely exploratory in nature, designed to provide preliminary information on how extremist websites might be identified by a web-crawler and how they might be used by terrorist or extremist groups.

Past studies have also explored the presence of extremist groups on the Internet and developed tools to collect extremist websites [25] [3] [7] [17] [19] [20]. Some organizations, including SITE institute, the Anti-Terrorism Coalition, and the Middle East Media Research Institute (MEMRI) have used manual analysis techniques to collect and monitor extremist websites. The Artificial Intelligence Lab uses automated processes for collection building. The Dark Web project

has combined both manual and automated processes to build and analyze collections with the goal of combining the efficiency of automated techniques with the accuracy of manual ones. The TENE project seeks to extend past work on web-site-collection tools. It represents a return to automated processes for the purposes of efficiency; however, it also seeks to achieve a degree of accuracy similar to manual processes. This is done through the use of specific inclusion criteria within the web-crawler, such as keyword requirements. However, TENE is still in its early stages of development and is undergoing further modifications so that, among other things, it can properly differentiate between extremist and non-extremist websites.

To begin, the list of keywords will be refined to remove infrequently used words and to include other words that may capture a variety of extremist websites (e.g., Jihadic websites, neo-Nazi or white supremacist websites, eco-terrorist websites, southern separatist web-sites, etc.). The process of refining the keyword list will naturally require analysis that extends far beyond the initial examination of the two extremist websites for this project. In addition, including the keywords in a variety of languages (English, Arabic, French, etc.) is expected to improve the crawler's detection capacity. At present, TENE can identify and record the presence of images and videos; in the future, it may also be designed to recognize other website tools, such as chat rooms, forums, and donation options. Detection of interactive tools on websites is particularly important, as, for instance, [12] identifies a shift from websites of individual Jihadist groups to websites of pan-Jihadist forums.

A further refinement will involve the integration of semantics tool for the purpose of contextual analyses. Specific text mining and clustering techniques have been developed to uncover themes or topics (i.e., clusters of semantically related words) within sets of documents. [5] used some of these tools to develop a content-based social network analysis (CB-SN) that they applied to various research and scientific networks. These tools include concept extraction and topic detection processes. Concept extraction identifies the relevant, domain-specific concepts after collecting textual information of interest (blogs, e-mails, articles, etc.). Topic detection uses a clustering algorithm to identify clusters of semantically similar concepts. In addition, researchers involved in the Dark Web project have begun exploring methods for sentiment and affect analysis in extremist websites [1]. Sentiment analysis distinguishes between text that contains positive and negative sentiments while affect analysis examines the emotions and moods expressed; in the Dark Web project, these analyses are achieved by looking at certain syntactic and stylistic features of websites [1]. By merging available tools with the web-crawler, emergent semantics and tone of language in extremist and terrorist websites can be more easily identified and information on popular topics can be extracted.

One of the ancillary benefits of TENE is its sustainability. This tool has already demonstrated its benefits in relation to the investigation of child exploitation on the Internet. Following its adaptation to the study of extremism, it will lend itself to further substantive applications, such as the funding of terror-

ism and the spread of propaganda. We also envision TENE as an integral part of a broader strategy to disrupt extremist networks.

References

1. Chen, W. (2012). Dark web: Exploring and data mining the dark side of the web. Springer: NY.
2. Chen, H., Chung, W., Qin, J., Reid, E., Sageman, M., & Weinmann, G. (2008). Uncovering the dark web: A case study of Jihad on the Web. Journal of American Society for Information Science and Technology, 59(8), 13471359.
3. Chen, H., Qin, J., Reid, E., Chung, W., Zhou, Y., Xi, W., Lai, G., Elhourani, T., Bonillas, A., Wang, F.-Y., & Sageman, M. (2004). The dark web portal: Collecting and analyzing the presence of domestic and international terrorist groups on the web. Proceedings of the sev-enth Annual IEEE Conference on Intelligent Transportation Systems.
4. Conway, M. (2002). Reality bites: Cyberterrorism and terrorist use' of the Internet. Retrieved from http://www.firstmonday.org/Issues/issue7_11/conway/index.html
5. Cucchiarelli, A., D'Antonio, F., & Velardi P. (2012). Semantically interconnected social networks. Social Network Analysis and Mining, 2(1), 69-95.
6. Frank, R., Westlake, B., & Bouchard, M. (2010). The structure and content of online child ex-ploitation networks. Proceedings of the tenth ACM SIGKDD Workshop on Intelligence and Security Informatics '04.
7. Institute for Security Technology Studies, Technical Analysis Group. (2004). Examining the Cyber Capabilities of Islamic Terrorist Groups. Retrieved from http://www.ists.dartmouth.edu/
8. Jacobson, M. (2010). Terrorist financing and the Internet. Studies in Conflict & Terrorism, 33(4), 353-363.
9. Jiang, C., Coenen, F., Sanderson, R., & Zito, M. (2009). Text classification using graph mining based feature extraction. Proceedings of the SGAI '09: International Conference on Artificial Intelligence. London.
10. Joffres, K., Bouchard, M., Frank, R., & Westlake, B. (2011). Strategies to disrupt online child pornography networks. Proceedings of the eleventh ACM SIGKDD Workshop on Intelli-gence and Security Informatics.
11. Kamruzzaman, S.M., Farhana, H., & Ahmen, R.H. (2009/10). Text classification using data mining. Proceedings of International Conference on Information and Communication Technology in Management '05. Multimedia University, Malaysia.
12. Musawi, M.A. (2010). Cheering for Osama: How Jihadists use Internet discussion forums. Quillan Foundation. Retrieved from http://www.quilliamfoundation.org/images/stories /pdfs/cheering-for-osama.pdf
13. National Research Council. (2002). Making the nation safer: the role of science and technology in countering terrorism. Washington, D.C.: National Academy of Sciences.
14. Office of Homeland Security. (2002). National strategy for homeland security. Washington D.C.: Office of Homeland Security.
15. Sebastiani, F. (2002). Machine learning in automated text categorization. ACM Computing Sur-veys, 34(1), 1-47.
16. Technical Analysis Group. (2004). Examining the Cyber Capabilities of Islamic Terrorist Groups. Hanover, NH: Institute for Security Technology Studies at Dartmouth College.

17. Thomas, T.L. (2003). Al Qaeda and the Internet: The danger of Cyberplanning'. Parameters, 33, 112-123.
18. Tong, S., & Koller, D. (2002). Support vector machine active learning with applications to text classification. Journal of Machine Learning Research, 2, 45-66.
19. Tsfati, W., & Weimann, G. (2002). www.terrorism.com: Terror on the Internet. Studies in Conflict & Terrorism, 25(3), 317-332.
20. Weimann, G. (2004). www.terror.net: How modern terrorism uses the Internet. Special Report, US Institute of Peace. Retrieved from http://www.usip.org/pubs/specialreports/sr116.pdf
21. Weimann, G. (2006). Terror on the Internet: The new Arena, The New Challenges. Washington, D.C.: United States Institute of Peace.
22. Westlake, B., Bouchard, M., & Frank, R. (2011). Finding the Key Players in Online Child Ex-ploitation Networks. Policy & Internet, 3(2), 104.
23. Xu, J., Chen, H., Zhou, Y., & Qin, J. (2006). On the topology of the Dark Web of terrorist groups. Lecture Notes in Computer Sciences, 3975, 367-376.
24. Zhou, Y., Qin, J., Lai, G., Reid, E., & Chen, H. (2005) Building knowledge management system for researching terrorist groups on the Web. Proceedings of the Eleventh Americas Conference on Information Systems '05. Omaha, NE, USA.
25. Zhou, Y., Reid, E., Qin, J., Chen, H., & Lai, G. (2005). U.S. extremist groups on the web: Link and content analysis. IEEE Intelligent Systems, 20(5), 44-51.